# 9 SAMPLE SIZE

## 9.1 Sample Size and Decision Errors

To determine the number of samples to collect, acceptable values of the Type I error rate ($\alpha$) and Type II error rate ($\beta$) must be specified as part of a statistical test. The process for doing this was discussed in Section 3.7. If there are many survey units and each unit requires a separate decision, even if $\mathbf{H_0}$ is true approximately 100$\alpha$% of the times the test is conducted, the null hypothesis will be incorrectly rejected. If a smaller value of $\alpha$ is used, the number of times this can be expected to happen decreases proportionately. On the other hand, larger values of $\alpha$ will reduce the number of samples initially required from each survey unit.

The power ($1-\beta$) is the ability of a statistical test to detect when the null hypothesis is indeed false and should be rejected. A test should have high power, i.e., small $\beta$, but smaller specified values of $\beta$ require a larger number of measurements.

The number of samples depends not only on $\alpha$ and $\beta$, but also on the width of the gray region relative to the measurement variability, $\Delta/\sigma$. This parameter essentially describes the resolution of the decision problem. When the resolution is high, only a few measurements are needed. When the resolution is low, many more measurements may be required, even though the specified $\alpha$ and $\beta$ are unchanged.

The DQO steps described in Sections 3.7 and 3.8 are used to balance the cost of sampling against the risks involved with any potential decision errors. It is important to realize, however, that the value of $\alpha$ is fixed at the desired value when the critical value for the test statistic is determined from the provided tables and used in the test. If the sample size is larger (or smaller) than planned, the effect will be an increase (or decrease) in the power, $1-\beta$. Similarly, if the measurement standard deviation is larger (or smaller) than anticipated, the effect will be a decrease (or increase) in the power.

The consequences of increasing or decreasing the power depends on whether Scenario A or Scenario B is being used. In Scenario A, where the null hypothesis is that the survey unit does not meet the release criterion, high power means that a survey unit that meets the release criterion has a high probability of passing the test. In Scenario B, where the null hypothesis is that the survey unit meets the release criterion, high power means that a survey unit that does not meet the release criterion has a high probability of failing the test.

In most cases, the sample sizes required can be determined using Tables 3.2 and 3.3. In the following sections, the assumptions made and the calculations performed in creating these tables are described. Methods for modifying the calculations under alternative assumptions are also given. It must be emphasized that relatively little effort is required to perform the suggested sample size determinations compared to the time and expense involved in collecting and analyzing samples. This is a key advantage to using the DQO process to determine sample sizes.

## 9.2 Sample Size Calculation for the Sign Test Under Scenario A

For the Sign test, the number of samples, $N$, required from the survey unit can be approximated from a formula given by Noether (1987):

$$N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{4(p-0.5)^2} \qquad (9\text{-}1)$$

where:

$\alpha$ = specified Type I error rate
$\beta$ = specified Type II error rate
$Z_{1-\alpha}$ = 100(1- $\alpha$) percentile of the normal distribution
$Z_{1-\beta}$ = 100(1- $\beta$) percentile of the normal distribution
$p$ = estimated probability that a random measurement from the survey unit will be less than the $DCGL_W$ when the survey unit median is actually at the LBGR. $p \neq 0.5$

Commonly used values for $\alpha$ and $\beta$, and the corresponding values of $Z_{1-\alpha}$ (or $Z_{1-\beta}$) may be found from Table 9.1. Other values can be obtained using any table of the cumulative standard normal distribution function, such as that in Appendix A.

**Table 9.1 Some Values of $Z_{1-\alpha}$ and $Z_{1-\beta}$ Used To Calculate the Sample Sizes**

| $\alpha$ (or $\beta$) | $Z_{1-\alpha}$ (or $Z_{1-\beta}$) |
|---|---|
| 0.01 | 2.3268 |
| 0.025 | 1.9604 |
| 0.05 | 1.6452 |
| 0.10 | 1.2817 |
| 0.20 | 0.8415 |

The numerator of Equation 9.1, $(Z_{1-\alpha} + Z_{1-\beta})^2$, depends on $\alpha$ and $\beta$, but not on $\Delta$ or $\sigma$. In addition, it only depends on the pair of values used for $\alpha$ and $\beta$, and is the same if these values are reversed. This can be seen in the symmetry of Table 9.2, where the value of $(Z_{1-\alpha} + Z_{1-\beta})^2$ is given for each pair of the values of $\alpha$ and $\beta$ listed in Table 9.1. As will be seen, these are also the minimum samples sizes for each pair of $\alpha$ and $\beta$ values.

The denominator of Equation 9.1 depends on the parameter $p$, but not on $\alpha$ or $\beta$. The definition of the parameter $p$ states that it is the estimated probability that a random measurement from the survey unit will be less than the $DCGL_W$ when the survey unit median is actually at the LBGR.

This is illustrated in Figure 9.1. The value of $1 - p$ expresses the likelihood that measurements exceeding the $DCGL_W$ will be observed, even if half of the concentration distribution is below the LBGR. This likelihood is higher when the measurement standard deviation is large compared to the width of the gray region. Some assumptions about the data distribution have to be made in estimating $p$. If it were possible to specify $p$ exactly, there would be no need to do the survey.

**Table 9.2 Some Values of $(Z_{1-\alpha} + Z_{1-\beta})^2$ Used To Calculate Sample Sizes**

| β | α 0.01 | 0.025 | 0.05 | 0.1 | 0.2 |
|---|---|---|---|---|---|
| **0.01** | 22 | 19 | 16 | 14 | 11 |
| **0.025** | 19 | 16 | 13 | 11 | 8 |
| **0.05** | 16 | 13 | 11 | 9 | 7 |
| **0.1** | 14 | 11 | 9 | 7 | 5 |
| **0.2** | 11 | 8 | 7 | 5 | 3 |

The relative width of the gray region, Δ/σ, is especially useful for estimating the parameter $p$. If the data are even approximately normally distributed, then $p$ can be estimated from

$$
\begin{aligned}
p &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{DCGL_W} e^{-(x-LBGR)^2/2\sigma^2}dx \\
&= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{LBGR+\Delta} e^{-(x-LBGR)^2/2\sigma^2}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\Delta}{\sigma}} e^{-x^2/2}dx \\
&= \Phi\left(\frac{\Delta}{\sigma}\right)
\end{aligned}
\qquad (9\text{-}2)
$$

Values of $p$ as a function of Δ/σ, computed from Equation 9-2, can be found in Table 9.3, or in the table of the cumulative normal distribution (Appendix A, Table A.1).

The factor $1/[4(p - 0.5)^2]$ in Equation 9-1 can be viewed as a multiplier applied to the sample sizes given in Table 9.2. If $p = 1$, this factor is one. Figure 9.2 shows the dependence of both $p$ and the sample size multiplier on the resolution of the decision problem, Δ/σ. Increasing Δ/σ beyond about three has little effect on reducing the sample size multiplier. Decreasing Δ/σ below about one causes the sample size multiplier to rise dramatically. The range from the smallest (3) to largest (22) sample size in Table 9.2 is about a factor of seven.  When Δ/σ is less than about 0.5, the sample size multiplier exceeds seven. Thus, small values of Δ/σ can have a bigger impact on increasing the sample size than the choice of acceptable decision error rates.

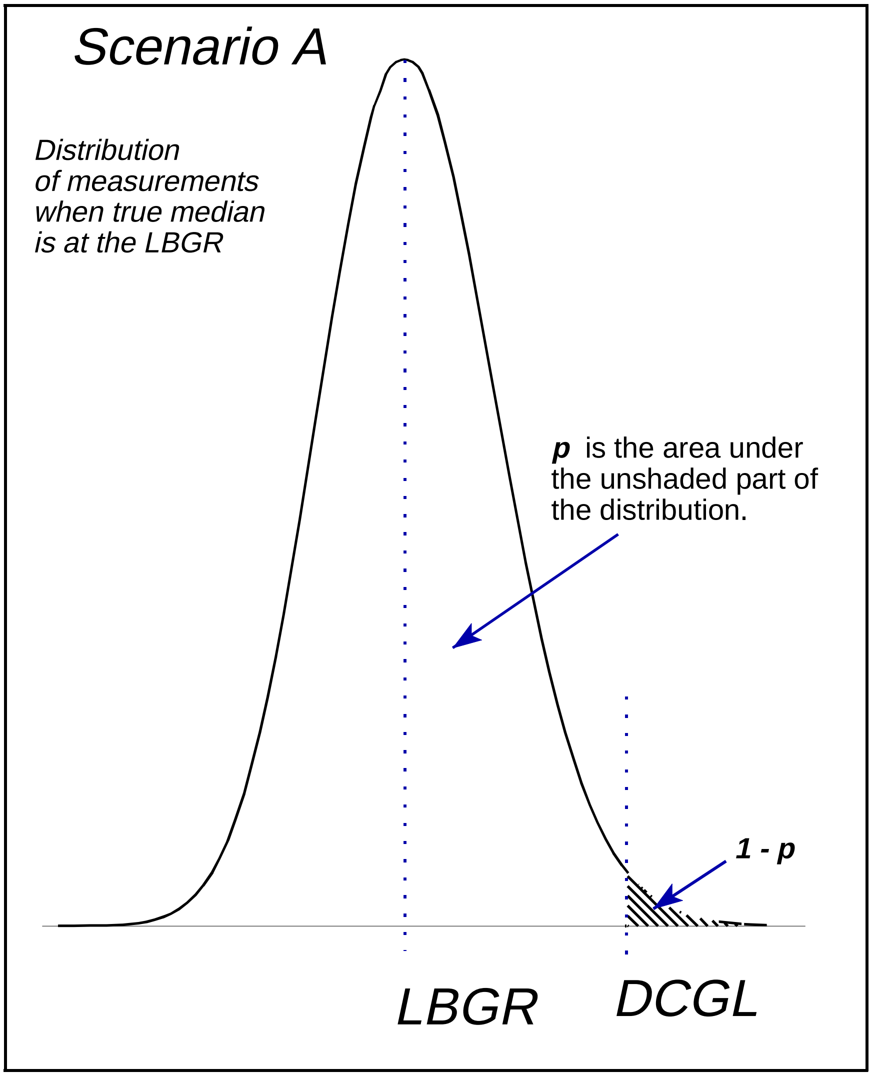


**Figure 9.1 The Parameter *p* for the Sign Test Under Scenario A**

The assumption of normality is not critical in the above calculations, since it is only being used to estimate an efficient sample size. However, if a different distribution is considered more appropriate, it can be used. Values of $p$ for other probability distributions with density function $f(x)$, mean equal to the LBGR, and standard deviation σ, can be computed from

$$p = \int_{-\infty}^{DCGL_W} f(x)\,dx \tag{9-3}$$

**Table 9.3 Values of *p* for Use in Computing Sample Size for the Sign Test**

| Δ/σ | *p* | Δ/σ | *p* | Δ/σ | *p* | Δ/σ | *p* |
|---|---|---|---|---|---|---|---|
| 0.1 | 0.53983 | 1.1 | 0.86433 | 2.1 | 0.98214 | 3.1 | 0.99903 |
| 0.2 | 0.57926 | 1.2 | 0.88493 | 2.2 | 0.9861 | 3.2 | 0.99931 |
| 0.3 | 0.61791 | 1.3 | 0.9032 | 2.3 | 0.98928 | 3.3 | 0.99952 |
| 0.4 | 0.65542 | 1.4 | 0.91924 | 2.4 | 0.9918 | 3.4 | 0.99966 |
| 0.5 | 0.69146 | 1.5 | 0.93319 | 2.5 | 0.99379 | 3.5 | 0.99977 |
| 0.6 | 0.72575 | 1.6 | 0.9452 | 2.6 | 0.99534 | 4.0 | 0.99997 |
| 0.7 | 0.75804 | 1.7 | 0.95544 | 2.7 | 0.99653 | 5.0 | 1.00000 |
| 0.8 | 0.78815 | 1.8 | 0.96407 | 2.8 | 0.99745 | | |
| 0.9 | 0.81594 | 1.9 | 0.97128 | 2.9 | 0.99813 | | |
| 1.0 | 0.84135 | 2.0 | 0.97725 | 3.0 | 0.99865 | | |

In some situations, it may be possible to estimate *p* directly from remediation control survey data, since it is simply an estimate of the proportion of the final status survey measurements that are expected to fall below the $DCGL_W$ .

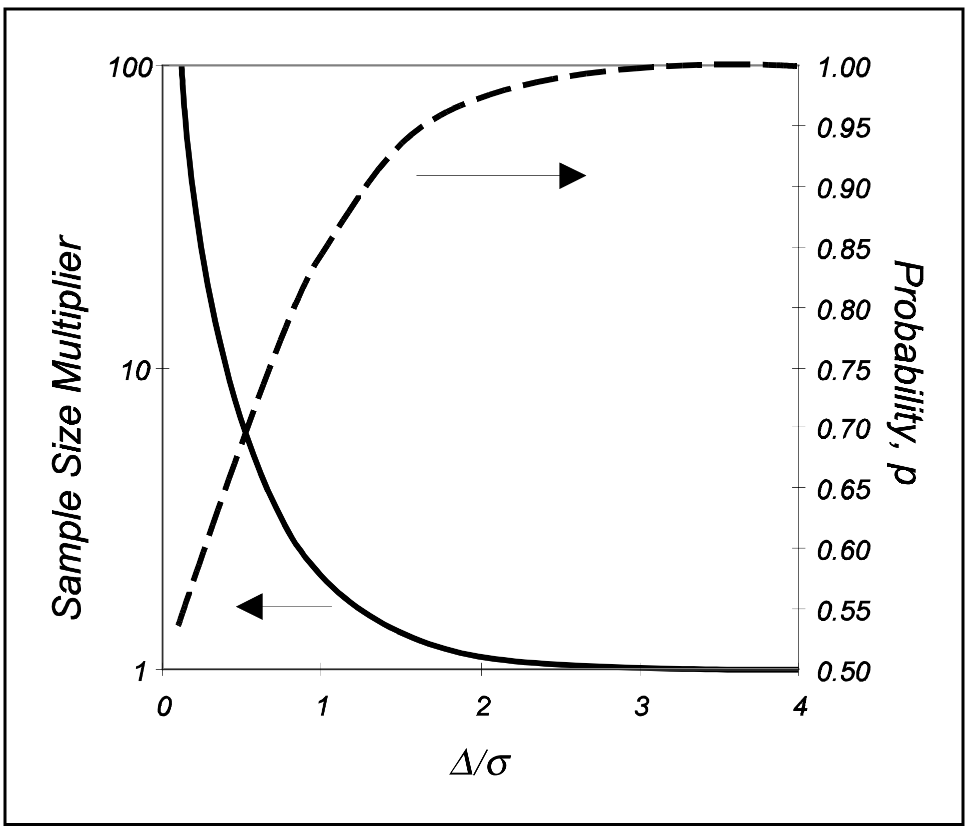


**Figure 9.2 Sample Size Multiplier and the Parameter *p* Versus Δ/σ**

One can also calculate $p$ from the estimated odds that a random measurement is less than the $DCGL_W$ versus that it is above the $DCGL_W$. If these odds are $r_1{:}r_2$, then $p = r_1/(r_1 + r_2)$. For example, if the odds that a random measurement is less than the $DCGL_W$ are 3:2, then $p = 3/(3+2) = 3/5 = 0.6$.

Whatever method is used to estimate p, it is important not to overestimate it, since that will result in a sample size inadequate to achieve the desired power of the test. The dependence of the sample size multiplier, $1/[4(p - 0.5)^2]$, on $p$ is shown in Figure 9.3.

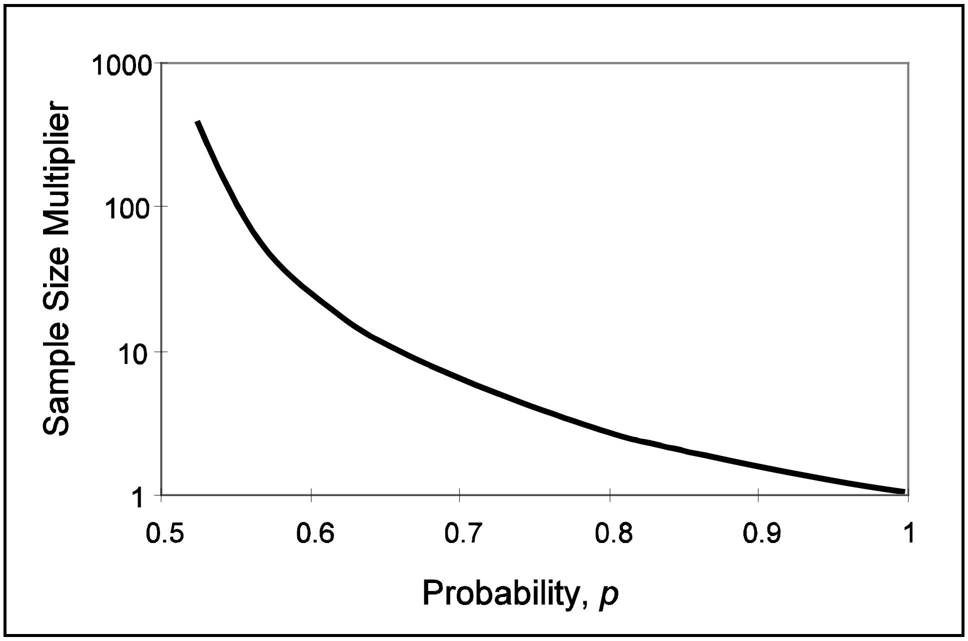


**Figure 9.3 Dependence of Sample Size on *p***

As an illustration, consider the example given in Section 5.1. For that example, the $DCGL_W = 15.9$, the LBGR = 11.5, $\alpha = \beta = 0.05$, and $\sigma = 3.3$. From Table 9.2, $(Z_{1-\alpha} + Z_{1-\beta})^2 = 11$ when $\alpha = \beta = 0.05$. This is the minimum sample size required for those values of the acceptable error rates. The width of the gray region, $\Delta = DCGL_W - LBGR = 15.9 - 11.5 = 4.4$, so $\Delta/\sigma = 4.4/3.3 = 1.3$.

From Table 9.3, the value of $p$ using the normal approximation is 0.903199. Thus the factor

$$
\begin{aligned}
1/[4(p - 0.5)^2] &= 1/[4(0.903199 - 0.5)^2] \\
&= 1/[4(0.403199)^2] \\
&= 1/[4(0.162569)] \\
&= 1/0.650278 \\
&\approx 1.54
\end{aligned}
$$

So, the minimum sample size of 11 is increased by a factor of 1.54 to 16.9. This would normally be rounded up to 17. However, because Equation 9-1 is an approximation, it is prudent to increase this number moderately. An increase of 20% is recommended. This increases the number of samples to 1.2(16.9) = 20.28, which is rounded up to 21. This is the number that appears in Table 3.2.

The effect of increased variability in the measurement data will be an increase in the required sample sizes. As $\Delta/\sigma$ becomes smaller, $p$ also becomes smaller. This decreases the denominator of Equation 9-1, increasing the sample size $N$ accordingly.

### 9.3 Sample Size Calculation for the Sign Test Under Scenario B

Under Scenario B, Equation 9-1 is also used to estimate the required sample size. The roles of $\alpha$ and $\beta$ are reversed, but this has no effect on the numerator of Equation 9-1, so Table 9.2 may still be used. The form of the denominator also remains the same, and Figure 9.3 still represents the dependence of the sample size multiplier on $p$. However, the definition of the parameter $p$ is different. The definition of the parameter $p$ under Scenario B is the estimated probability that a random measurement from the survey unit will be greater than the LBGR when the survey unit median is actually at the $DCGL_W$. This is illustrated in Figure 9.4. The value of $1-p$ expresses the likelihood that measurements less than the LBGR will be observed, even if half of the concentration distribution is above the $DCGL_W$. This likelihood is higher when the measurement standard deviation is large compared to the width of the gray region.

If, as in Scenario A, we assume that the data are approximately normally distributed, the width of the gray region, $\Delta/\sigma = (DCGL_W - LBGR)/\sigma$, can be used to estimate the parameter $p$:

$$
\begin{aligned}
p &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{LBGR}^{\infty} e^{-(x-DCGL_W)^2/2\sigma^2} dx \\
&= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{(LBGR-DCGL_W)+DCGL_W}^{-\infty} e^{-(x-DCGL_W)^2/2\sigma^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{LBGR-DCGL_W}{\sigma}}^{\infty} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{DCGL_W-LBGR}{\sigma}} e^{-x^2/2} dx \\
&= \Phi\left(\frac{\Delta}{\sigma}\right)
\end{aligned}
\tag{9-4}
$$

This is the same as Equation 9-2. Even though the definition of $p$ has changed, its value as a function of $\Delta/\sigma$ has not changed.

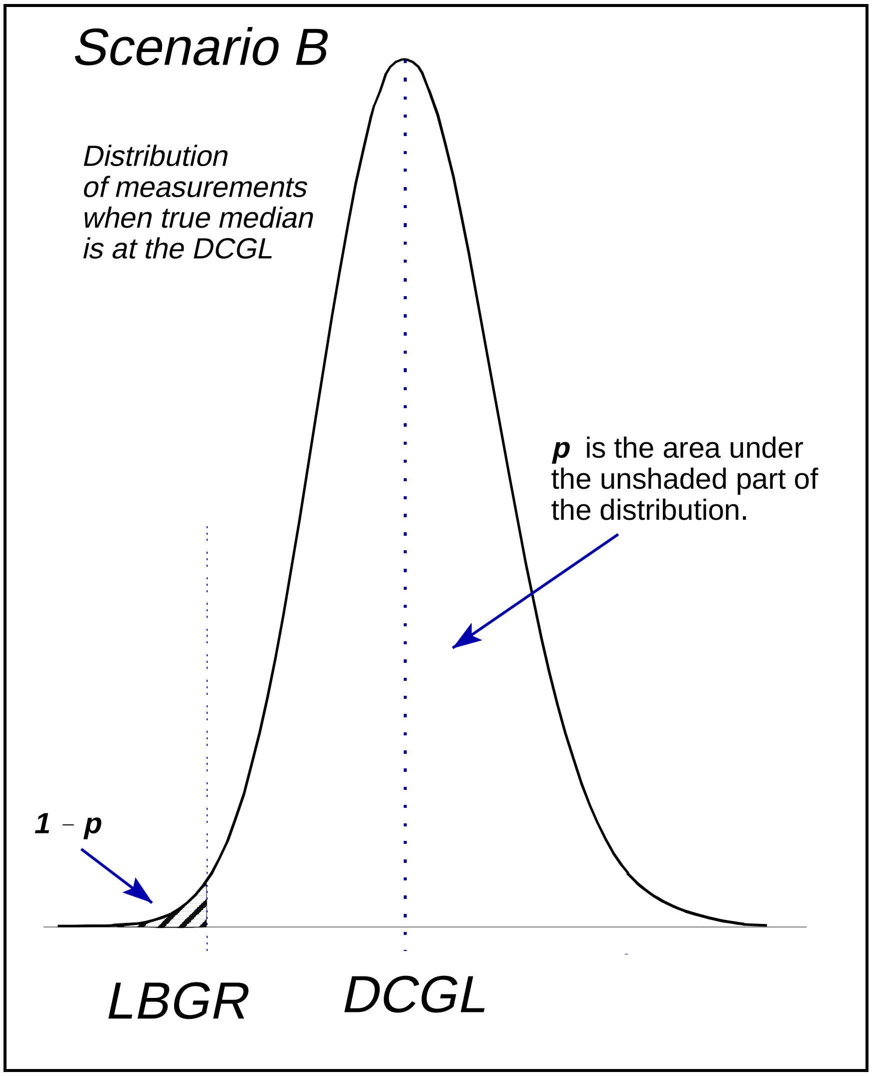


**Figure 9.4 The Parameter *p* for the Sign Test Under Scenario B**

The above calculation shows that the value of $p$ computed from Equation 9-2, as found in Table 9.3, or in the table of the cumulative normal distribution (Appendix A) can be used in Scenario B as well as Scenario A. Figure 9.2, expressing the dependence of sample size on $\Delta/\sigma$ is unchanged, and that is why only one version of Table 3.2 is needed for both scenarios.

If a distribution other than normal is considered more appropriate for determining $p$, the following equation can be used. For a probability distribution with density function $f(x)$, mean at the $DCGL_W$ , and standard deviation $\sigma$,

$$p = \int_{LBGR}^{\infty} f(x)\,dx \qquad (9\text{-}5)$$